Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Modelling Logistic Growth Model for COVID-19 Pandemic in India

Minni Jain
Delhi Technological University
Department of CSE
New Delhi, India
minnijain@dtu.ac.in

Prajwal Kumar Bhati
Delhi Technological University
Department of CSE
New Delhi, India
prajwalroxman@gmail.com

Pranav Kataria
Delhi Technological University
Department of CSE
New Delhi, India
pranavkataria_bt2k16@dtu.ac.in

Rohit Kumar
Delhi Technological University
Department of CSE
New Delhi, India
rohitmamc2016@gmail.com

Abstract-- **An early analysis of growth dynamics for infectious diseases, like COVID-19, is needed to dissect the crucial driving factors that result in rapid disease transmission, refine the measures taken to control the pandemic and improve disease forecast. The phenomenological models are used to identify the initial climbing growth period of COVID-19 outbreak in India and have modelled 3 major epidemic growth models: Generalized logistic growth, Logistic growth and Generalized growth, to predict the growth in the total number of positive cases, daily increase in the number of positive tested cases and the daily growth rate in confirmed positive cases, dated from Apr 10, 2020, to Apr 20, 2020. The bootstrap resampling method is applied for data prediction to process the sample data, dated from Jan 31, 2020, to Apr 10, 2020, and to calculate the 3 major growth parameters: r (Rate of growth at an early stage), K (Final epidemic size) and C(Number of aggregate cases at time t), which are used to calculate confidence intervals which predict the future direction of the curve and increase in the number of confirmed cases with 95% accuracy for the interval Apr 10, 2020, to Apr 20, 2020. Our models predict exponential and sub-exponential spread rate in the number of positive cases in India from Apr 10, 2020, to Apr 20, 2020. Our findings reveal that significant measures are needed to control the transmission rate of the virus in the community, as the models predict sub-exponential growth in India.**

Keywords – COVID-19; Phenomenological models; Epidemic Growth Models; Logistic Growth; Bootstrap Resampling Method; Growth parameters; Confidence Intervals; Data Prediction.

I. INTRODUCTION

In the early stages of an epidemic, the growth rate of the infection spread follows an exponential path. Extrapolation of available data is the most crucial method used to depict the growth dynamics of the epidemic. The predicted data help the policymakers to take preventive measures in the right direction to curb the spread of disease in the community.[11] In case of an epidemic, epidemic growth models are constructed to extrapolate the rate at which the disease might spread in the community. These epidemic growth models are constructed using the parameters obtained by exploring available data, using bootstrap resampling methods. In the bootstrap resampling method, multiple simulations run on the available dataset. In consecutive simulations, parameters are adjusted to extrapolate the most precise estimates that are further used to plot the respective epidemic growth models.[12] COVID-19 disease is now proclaimed as a pandemic around the globe. India comes among one of the few countries in the world which have implemented complete lockdown as a preventive measure to decrease the growth in positive cases of COVID-19 disease at an early stage of an outbreak. It has now spread across the whole world, badly affecting nations like the USA, Italy and Spain. This paper focuses on extrapolation of available data for India to construct early-stage epidemic growth models to dissect the crucial driving factors that are resulting in the rapid spread of disease, to refine the measures taken by authorities to control the pandemic and improve disease forecast.[4,5].

Logistic growth model construction is favourable and can unravel the growth and fall in the number of cases around the inflection point. If the logistic curve converges at an early stage, it means the pandemic is under control, suggesting that the steps taken to control the pandemic are working. Whereas, if the logistic curve does not converge and shows an exponential growth even after the occurrence of the inflection point, then it indicates that the measures taken in order to manage the pandemic are not working, and the epidemic is still in its early growth stage. Logistic models also involve calculation of important feature parameters like rate of growth at early stage(r), final epidemic size (K) and the number of aggregate cases at time t (C), growth deceleration (p), disease turning point(t) which helps in accurate modelling of curve. Hence, the proposed method of logistic modelling is meritorious over conventional methods.

I. LITERATURE SURVEY

Novel coronavirus (COVID-19) belongs to a family of coronaviruses, which can cause various illness like
SARS and MERS.[9] According to the World Health Organization, COVID-19 first appeared in the city of Wuhan, Hebei province, in China in December 2019. SARS-CoV-2 is a new mutation Coronavirus, which is identified as the cause of coronavirus disease of 2019.[8]This pandemic has now affected almost every country on the planet, and researchers around the world are racing to find a cure (or vaccine) for this virus.[9] USA, Italy and Spain are the worst hit nations.The situation in USA is very severe, because more than 50,000 people have lost their lives to this virus.[8] Active research in the field of Computers and Mathematics has helped researchers and doctors, to track the rate of growth of this virus in different countries.[1] Research referenced in [1] has shown that the rate of growth of virus in every country follows the same pattern, that is exponential in nature.

II. METHODS AND THEORY

A. DATASET SOURCES

This report is focused on daily data for the number of positive recorded cases from India. The dataset is sourced from the COVID-19 Patient Database[4] which consists of total confirmed cases state wise consisting of patient number, patient gender, age-bracket, travel history,detected city and current status for all cities in India. The data is updated daily. In this report the data from Jan 31 is used when the first case in India was reported to Apr 10 and used the models to forecast outbreak for next 5 to 10 days.

B. EPIDEMIC GROWTH MODELS

India is currently at the initial stage of the spread, that is, the number of positive cases has not reached a saturation peak yet, and the daily increase is still widening.
In the present scenario, the general epidemic growth models like Richards model and complete interval epidemic curves which constitute an extra variable ‘α’ signifying the imbalance between rise and sink for the growth frequency curve, cannot be constructed as the amount of data is very limited.
Early epidemic growth models usually result in an ‘S’ shape when they reach saturation stage. This happens when the epidemic is at an early stage and the data available for processing the model is limited.
Hence, this research focuses on 3 major models, which are constructed using few parameters:

1. Logistic Growth Model

$$\frac{dC(t)}{dt} = r(1 - \frac{C(t))}{K})C(t) \quad (1)$$

2. Generalized Logistic Growth Model(GLM):

$$\frac{dC(t)}{dt} = r(1 - \frac{C(t))}{K})C^p(t) \quad (2)$$

3. Generalized Growth Model (GGM):

$$\frac{dC(t)}{dt} = rC^p(t) \quad (3)$$

Here, the following main parameters are calculated to construct accurate growth models:

1. R : Rate of growth at early stage.
2. K : Final epidemic size.
3. C(t) : Number of aggregate cases at time t.
4. p : Signify profile of distinct growth.

Parameter p lies from 0 to 1, where if:

a. $p = 0$: Implies constant growth rate.
b. $0 < p < 1$: Implies sub-exponential growth rate.
c. $p = 1$: Implies exponential growth rate.

The 2 basic models, which are derived from Richard’s model keeping the parameter α equal to 1 (α=1) and p=1, result in logistic development and generalized logistic development models. The development rate for these models deteriorates proportionally to the number of days. These curves tend to saturate the development in the figures for positive cases at an early juncture of epidemic resulting in optimistic results.
The (3) generalized growth model (GGM) allows for a steady increase in outbreak in the early phase (p<1), but it is unable to explain the decomposition of the frequency rate at advanced stages. Therefore, it acts as an upper barrier, or the worst case scenario that is obtained on the presumption that the upsurge will continue to develop following an expanding pattern as before.Logistic development (1) and Generalized Logistic development (2) usually result in an ‘S’ shaped curve, when the growth in the number of confirmed cases saturates. They predict very optimistic results, which can be seen as the best case scenario in the early stage analysis of the pandemic.

The generalized logistic growth model (2) results in sub-exponential development rate at a preliminary phase, hence it is preferred to study the imbalance between decay and growth dynamics. These three models are constructed using the available data for the total number of actual recorded cases in India.

C. PARAMETER ESTIMATION

For the parameter estimation stage, the Damped least square method, also known as Marquardt-Levenberg algorithm has been used to resolve the problem of least square nonlinear regression for curve fitting. In order to fit the normal function of normal growth in generalized case of Richard’s model (α =1, p = 1), commence the starting spike and permit the algorithm to calculate one of the 3 parameters optimally. Since, the early phase of growth models does not observe the logistic rate of development, hence to calculate the variability in our model variables, bootstrap resampling method along with a Poisson function is utilized for error estimation $f(\lambda)$, where λ is the expected value of the distribution as in prior inspections[2].

D. CONFIDENCE INTERVAL

The confidence interval mean is calculated for parameter estimation developed for Models (1), (2) and (3) by iterating 200 sets of realizations using parametric bootstrap resampling technique as used in prior studies[2].Parameters were then estimated from the realizations to construct 95% accurate Confidence Intervals. A Poisson error structure was used in the parametric bootstrap method for confidence interval estimation. Problems with the appropriate procedure occur when the model cannot be specified separately from the data resulting in unstable variations of the estimated parameters.The models parameters were evaluated for accurate identification using the simulations.

III. RESULTS

The Generalized linear model parameter p, which signifies the distinct profile growth of infected population is estimated to be 0.99 (95% Confidence Interval: [0.98, 1]). This high value of parameter p signifies that the growth rate is heading towards an exponential trend, and major preventive measures need to be enforced at the earliest.

On the contrary, the logistic model depicts a very optimistic best-case scenario , forecasting the total number infected population at 9,547 (95% Confidence Interval: [9349, 9746]) after 5 days (April 15), and 10,723 (95% Confidence Interval: [10514, 10933]) after 10 days.It might be a very optimistic result if the actual infected cases follow the logistic growth model. India announced their confirmed case for the first time on Jan 30. Since 10 April 2020, according to Indian administration, there were 206 deaths in India and 7,062 cases of infection, indicating that the mortality rate stands at 2.91% among the infected population.

While the growth rate of the infected population is slowing as predicted in Figure 5. , future conditions stay undetermined as the total number of infected individuals is close to the tip of inflection. Also, there is a possibility of new major outbursts to occur in the near future.

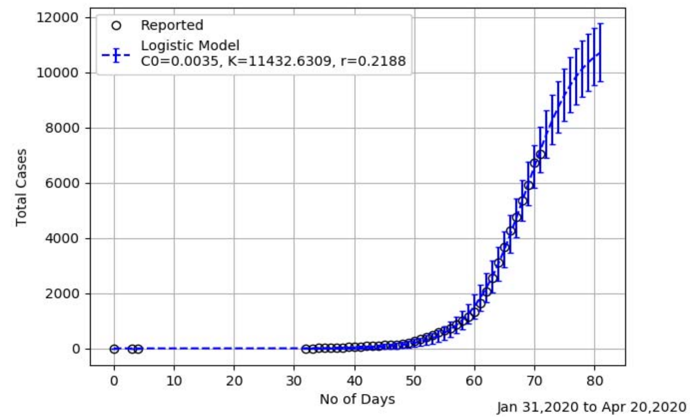

Figure 1. Estimation of Logistic Model to total cumulative confirmed cases in India since Jan31. The error bar is 95% CI extracted from 200 simulations using a poisson error structure.

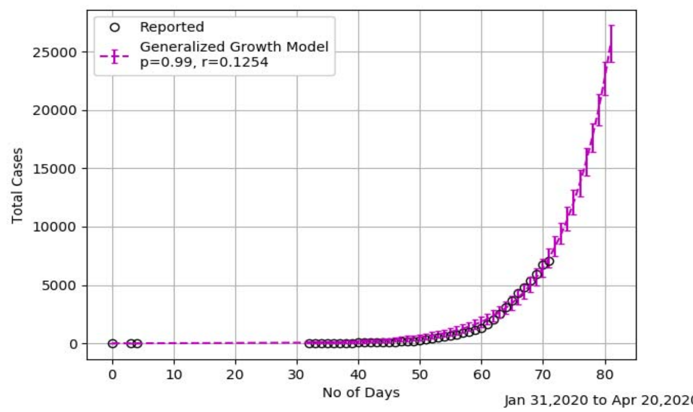

Figure 2. Estimation of Generalized Growth Model to total cumulative confirmed cases in India since Jan31. The error bar is 95% CI extracted from 200 simulations using a poisson error structure

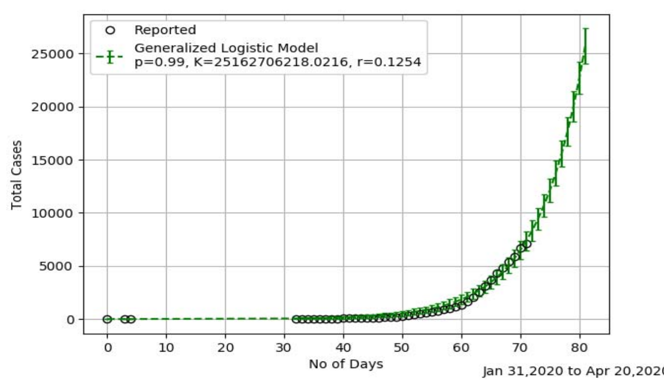

Figure 3. Estimation of Generalized Logistic Model in green to total cumulative confirmed cases in India since Jan31. The error bar is 95% CI extracted from 200 simulations using a poisson error structure.

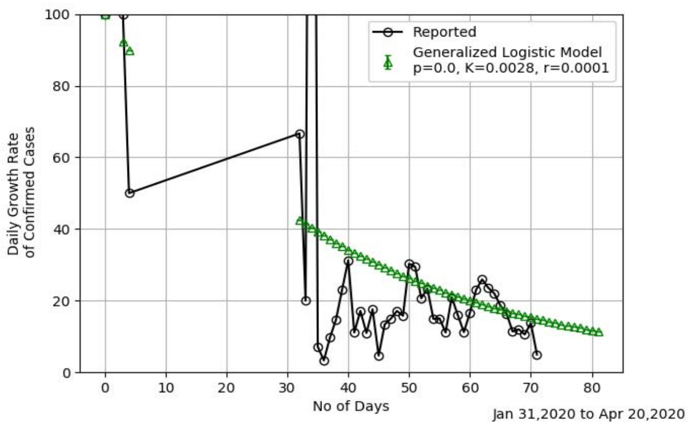

Figure 4. Estimation of models[3](*Generalized Logistic Development Model[1] (in green)*) to rate of daily development of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

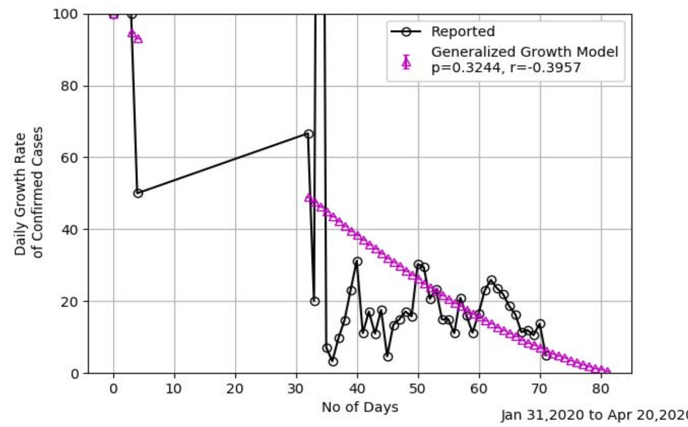

Figure 6. Estimation of models[3](*Logistic Development Model[1] (in blue*),Generalized Growth Model (*in pink*) to rate of daily development of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

The probability is that the growth rate can keep on fluctuating at current levels for sometime, where the generalized growth model tends to be more favorable for drawing inferences. In this case, the infected population will increase to 13735 (95% CI: [13499, 13970]) in 5 days (April 15*th*), also to 25706 (95% CI: [25377,26034] within next 10 days (April 20*th*).

Figure 7. predicts that the point of inflection was passed a few days back, and also the daily number of confirmed infections has started decreasing. This seems to be a positive indication suppose the disease to break out following the logistic depreciation in cases in the following days, presuming the absence of other major public transmission, as was witnessed during phase III of most hotspots in China's research [1].

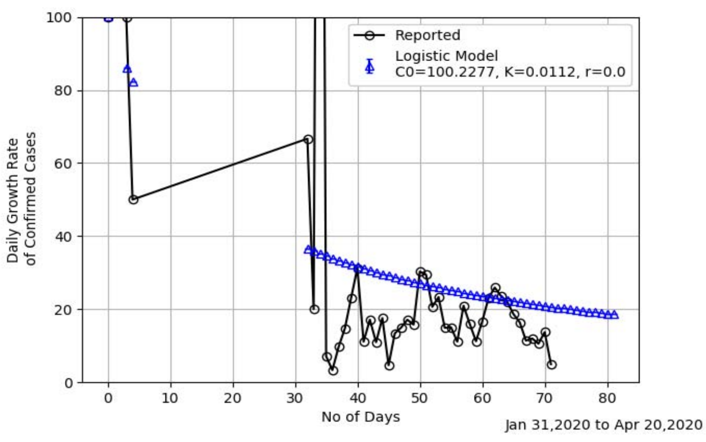

Figure 5. Estimation of models[3](*Logistic Development Model[1] (in blue*),Generalized Growth Model (*in pink*)to rate of daily development of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

The daily development in the infected population is likely to deteriorate from this junction with an estimated daily surge of 603 (95% Confidence Interval:[550,655]) in 10 days(Apr 20) as predicted by the Logistics Model.

A pessimistic result is obtained from the Generalized Logistic model and Generalized Growth model. Both models estimate the daily increase of infected cases to be : 1355 (95% Confidence Interval: [1279, 1431]) from GLM and 1780 (95% Confidence Interval: [1697,1862]) from GGM, both of which could be predicted to be attained within 10 days

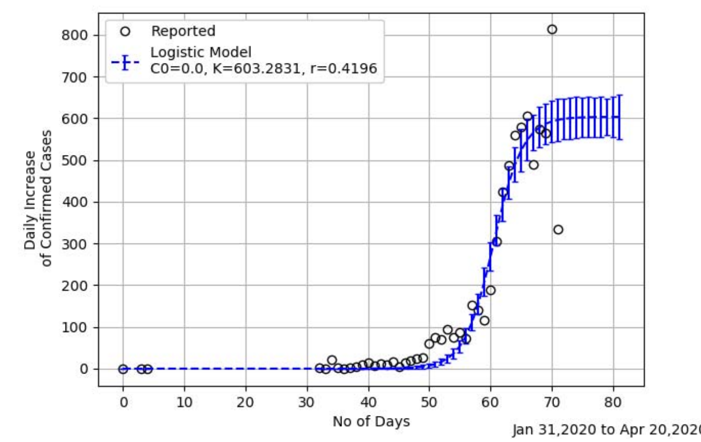

Figure 7. Estimation of models[3] (*Logistic Development Model [1] (in blue)*) to daily increase of infected population in India since Jan 31.The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

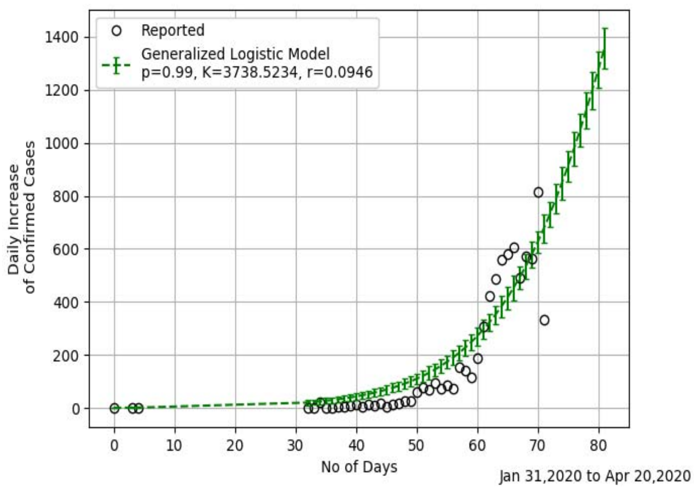

Figure 8. Estimation of models[3](*Generalized Logistic Development Model*[1] *(in green)*) to daily increase of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

As shown in Figure 4. and Figure 6. the daily growth rate of the confirmed positive cases is definitely not growing exponentially, because the growth rate reduced notably from 40% in February to 1.5% by April 10.The initial period of the Indian outbreak occurred on Jan 30, where most cases are admitted to have a clear travel history to infected places.Although, it is observed from the results that the surge in figures is approaching a constant trend, the uncertainty in future cases might occur.

IV. CONCLUSION AND FUTURE WORK

In this research paper, uniquely introduced variations of logistic models to data of covid-19 pandemic in India and have employed a special 3 parameter Generalized Logistic Model and 2-parameter Generalized Growth Model to model the spread of this outbreak in India. Levenberg–Marquardt algorithm is used for parameter estimation to carry-out least square nonlinear minimization for curve fitting in Python and have calculated confidence intervals with 95% average accuracy, using bootstrap resampling method , where 200 simulations are performed using Poisson error structure. For some days, the number of total cases is predicted with more than 97% of accuracy, which shows the reliability of the models constructed.

With this research paper, tried to model the following growth models: (1)logistic growth development model, the (2)generalized growth development model,(3)generalized logistic growth model to the actual infected population of Covid-19 pandemic from Jan. 30 to Apr 10 of India. Our analysis tries to discover the growth of the covid-19 pandemic in India and the influence of National Lockdown at the average level.

On the grounds of experience in China[1], a high level risk in India is predicted with approximate total actual cases on April 14 to be 12116 (i.e. 95 % CI: [11895, 12336]), 25706 (i.e. 95 % CI: [25337, 26034]) by April 20.The GLM parameter p is around 0.99 (95% Confidence Interval: [0.98, 1]). This indicates that domestic transmission sources in India are out of control as growth is nearing an exponential one.On the contrary, the best case scenario would be the one described in the logistic model, giving the total infected population as 9547 (95% CI: [9349, 9746]) after 5 days (April 15), and 10723 (95% CI: [10514, 10933]) after 10 days. It might be a very good result if the actual infected cases follow the logistic model.Extrapolation of data for India has helped to determine the course of future measures that need to be taken in order to curb the development in COVID-19 disease in India. National lockdown has definitely helped to curb the surge in positive cases in India, but the infection might start spreading rapidly if the lockdown is lifted without prior precautions.

Accuracy of results depends proportionally on the size of the sample data collected. In our research, modelling logistic growth models for COVID-19 pandemic in its most nascent early stages, inferring that the size of sample data available for data extrapolation is very small. Hence, as the sample data size is very limited, saw differences between actual and predicted forecasts using the proposed method.Some external factors also might have resulted in mis-reporting of actual cases.

Table I: Short Term Forecast for daily increase of confirmed cases as predicted by GLM along with actual data and accuracy(%)

Day	Date	Actual	Predicted	Accuracy (%)
72	11 April 2020	854	677	79.27
73	12 April 2020	758	730	96.30
74	13 April 2020	1243	787	63.31
75	14 April 2020	1031	847	82.15
76	15 April 2020	886	911	97.25
77	16 April 2020	1061	978	92.17
78	17 April 2020	922	1188	77.6
79	18 April 2020	1371	1268	92.48
80	19 April 2020	1580	1274	80.63

Under-testing due to unavailability of testing kits in India might have also resulted in mis-reporting in actual recorded cases, which further resulted in deviation from actual figures in India.

Also, new hotspots might have originated in most affected states like Maharashtra, which might have led to sudden disruptions in the number of reported cases in April 2020.

Regression convergence depends significantly on preliminary values of parameters K, r and p, it might fail to converge accurately for a limited sample dataset.

Therefore, logistic models are highly data-driven in nature and the accuracy in predicted data depends on how well can the data collected mimic the actual dynamics of the pandemic.

The logistic models constructed mimic both the spread of epidemic and also the control in epidemic (S-shaped curve). Curve starts to converge after inflection point, predicting the control in the epidemic. Non-convergence of logistic models implies that the pandemic is not in control and further measures need to be undertaken.

REFERENCES

[1] Ke Wu1,2, Didier Darcet3, Qian Wang4 and Didier Sornette. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. Preprint: https://www.medrxiv.org/content/10.1101/2020.03.11.20034363v1

[2] Chowell et al., 2007; Chowell et al., 2006a. Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. Preprint at: https://www.sciencedirect.com/science/article/abs/pii/S0022519305005096

[3] Junling Ma · Jonathan Dushoff ·Benjamin M. Bolker · David J.D. Earn. Estimating Initial Epidemic Growth Rates. Preprint at: https://link.springer.com/content/pdf/10.1007/s11538-013-9918-2.pdf

[4] COVID-19 Patient Database for India available at https://docs.google.com/spreadsheets/d/e/2PACX-1vSz8Qs1gE_IYpzlkFkCXGcL_BqR8hZieWVi-rphN1gfrO3H4lDtVZs4kd0C3P8Y9lhsT1rhoB-Q_cP4/pubhtml#

[5] A Literature Review of 2019 Novel Coronavirus During the Early Outbreak Period: Epidemiology, Causes, Clinical Manifestation and Diagnosis, Prevention and Control.

[6] Tian-Mu Chen, Jia Rui, Qiu-Peng Wang,Ze-Yu Zhao, Jing-An Cui, Ling Yin;A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Preprint at: **https://doi.org/10.1186/s40249-020-00640-3**

[7] Xinguang Chen, Bin Yu; First two months of the 2019 Coronavirus Disease (COVID-19) epidemic in China: real-time surveillance and evaluation with a second derivative model. https://doi.org/10.1186/s41256-020-00137-4

[8] Poudel Adhikari, S.; Meng, S.; Wu, Y.; Mao, Y.; Ye, R.; Wang, Q.; Sun, C.; Sylvia, S.; Rozelle, S.; Raat, H.; Zhou, H. A Literature review of 2019 Novel Coronavirus During the Early Outbreak Period: Epidemiology, Causes, Clinical Manifestation and Diagnosis, Prevention and Control. *Preprints* **2020**, 2020020060 https://www.preprints.org/manuscript/202002.0060/v2

[9] World Health Organization article(WHO); Naming the coronavirus disease (COVID-19) and the virus that causes it https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

[10] Cécile Viboud, Lone Simonsen, Gerardo Chowell; A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks https://doi.org/10.1016/j.epidem.2016.01.002

[11] Epidemiology article on wikipedia: https://en.wikipedia.org/wiki/Epidemiology

[12] Bootstrap resampling method; Medium Article https://towardsdatascience.com/an-introduction-to-the-bootstrap-method-58bcb51b4d60